Harnessing Organizationally Distributed Data with VeilFS


ACK Cyfronet AGH
Department of Computer Science, AGH UST

KUKDM 2014
Zakopane, Poland, March 12-14, 2014
Addressed problem

The user is typically interested in results - he expects that data access will be simple while data management in distributed environment is often too complicated for many users.

The access to the data in large scale computational environment should be the same as on personal computer.

Administrator is interested in effective utilization of resources - he expects that data management will be simple.

Data management in large distributed environments where many problems appear (several different technologies, accounting, users’ groups) should be automated and supported by unified tools.

How to do that?
Overview of VeilFS

Management of organizationally distributed data due to cooperation of independent instances of VeilFS
VeilFS – user point of view

Ease of use and unified access from anywhere

Mount VeilFS at your PC and use your data as if they were local (currently only Linux).

Or simply use Internet browser (all operating systems).

Process data at ANY worker node.
VeilFS – user point of view

- Ease of use and unified access from anywhere
- Data publication

Simple sharing in groups (throw data to group directory)

Ease and stop sharing if needed
VeilFS – user point of view

- Ease of use and unified access from anywhere
- Data publication and sharing within groups
- Metadata support

Easily manage metadata using Internet browser.

Or manage metadata using command line at any node.
Exemplary use case with 2 sites

User 1 processes data using services deployed in Site 1.

Users 1, 2, 3 belongs to a group that works together.

User 2 sees data produced by User 1 in any site.

Data can be streamed or migrated permanently on fly.

User 2 is able to process this data using services deployed in Site 2.

User 3 can analyze data produced by coworkers and publish chosen results using Web-based GUI.

Anyone is able to download chosen results.
Integration with other services

- VeilFS provides REST API
  - API allows direct interfacing from third party applications
  - third party applications can interact with VeilFS on behalf of the users with standard X509 proxy certificates

- Domain Grids – main client of REST API
  - VeilFS can provide data to computing elements where services are deployed
  - VeilFS can manage data storing and migration when services do not have access to shared storage

- REST API evolution – VeilFS integration with Grid middleware
  - VeilFS can provide information about data that helps job schedulers in choice of best site for the job
Preliminary tests results

Transfer rates measured by SysBench [Mb/s]

- 16 threads, RW
- 16 threads, WR
- 16 threads, RD
- 1 thread, RW
- 1 thread, WR
- 1 thread, RD

- Without VeilFS
- With VeilFS
Conclusions

- From user point of view the system:
  - is easy in use,
  - provides unified access to data,
  - supports publication and sharing of data,
  - **overcomes all existing barriers**.

- VeilFS has also second, not shown today face – it helps administrators during the data and infrastructure management.

- First ongoing deployment in PL-Grid base at one site resources:
  - multi-site deployment is planned after further tests.